



# An Approximate Bayesian Computation approach for inferring patterns of cultural evolutionary change

E.R. Crema <sup>a,\*</sup>, K. Edinborough <sup>a</sup>, T. Kerig <sup>a,b</sup>, S.J. Shennan <sup>a</sup>

<sup>a</sup> UCL Institute of Archaeology, 31–34 Gordon Square, London WC1H 0PY, UK

<sup>b</sup> Archaeology of Pre-Modern Economies (DFG Research Training Group 1878), Universities of Cologne and Bonn, Albertus-Magnus-Platz, D-50923 Koeln, Germany

## ARTICLE INFO

### Article history:

Received 13 March 2014

Received in revised form

9 July 2014

Accepted 12 July 2014

Available online 21 July 2014

### Keywords:

Cultural evolution

Unbiased transmission

Frequency-biased transmission

Agent-based simulation

Approximate Bayesian Computation

Equifinality

Neolithic Europe

## ABSTRACT

A wide range of theories and methods inspired from evolutionary biology have recently been used to investigate temporal changes in the frequency of archaeological material. Here we follow this research agenda and present a novel approach based on Approximate Bayesian Computation (ABC), which enables the evaluation of multiple competing evolutionary models formulated as computer simulations. This approach offers the opportunity to: 1) flexibly integrate archaeological biases derived from sampling and time averaging; 2) estimate model parameters in a probabilistic fashion, taking into account both prior knowledge and empirical data; and 3) shift from an hypothesis-testing to a model selection approach. We applied ABC to a chronologically fine-grained Western European Neolithic armature assemblage, comparing three possible candidate models of evolutionary change: 1) unbiased transmission; 2) conformist bias; and 3) anti-conformist bias. Results showed that unbiased and anti-conformist transmission models provide equally good explanatory models for the observed data, suggesting high levels of equifinality. We also examined whether the appearance of the Bell Beaker culture was correlated with marked changes in the frequency of different armature types. Comparisons between the empirical data and expectations generated from the simulation model did not show any evidence in support of this hypothesis and instead indicated lower than expected dissimilarity between assemblages dated before and after the emergence of the Bell Beaker culture.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

## 1. Introduction

In the last 30 years similarities between genetic and cultural evolutionary processes have fomented a rich history of cross-disciplinary studies bridging biological and social sciences, often under the umbrella-term of dual inheritance, or gene-culture coevolution theory (Cavalli-Sforza and Feldman, 1981; Boyd and Richerson, 1985; Henrich and McElreath, 2003). Archaeologists have not been immune to this cross-fertilisation, and their active commitment is testified by a variety of studies, including (but not limited to) reconstructions of cultural phylogenies (e.g. O'Brien and Lyman, 2003; Collard et al., 2006; Cochran and Lipo, 2010; O'Brien et al., 2014), analysis of dynamics in demic and cultural diffusion (e.g. Coward et al., 2008; Steele, 2009; Fort, 2012), modelling and analysis of subsistence (MacDonald, 1998; Richerson et al., 2001; Lake and Crema, 2012) and technological change (Bettinger and

Eerkens, 1999; Fitzhugh, 2001; Andersson, 2011), enquiries on social inequality (e.g. Bentley et al., 2005; Shennan, 2011b), and identification of patterns of social learning in relation to changes in demography (e.g. Shennan, 2001; Henrich, 2004; Conolly et al., 2008; Powell et al., 2009), as well as the properties (e.g. Eerkens and Lipo, 2007; Lycett and von Cramon-Taubadel in press) and frequency of cultural variants (e.g. Neiman, 1995; Shennan and Wilkinson, 2001; Kohler et al., 2004; Steele et al., 2010).

This paper follows this rich tradition of studies, focussing on one of the most long-lasting enquiries of our discipline: the patterns and the processes of temporal change in the frequency of cultural variants. Our starting point is the notion of “unbiased” cultural transmission (Boyd and Richerson, 1985), a model of social learning where the probability of adopting a cultural variant is a function of its relative frequency in the population. Whilst this implies that the frequency of each cultural variant remains unchanged over time, the random nature of this copying process, combined with episodes of cultural innovation where new variants are introduced in the population, determine fluctuations in the frequencies of each variant. Neiman's pioneer work (1995) highlighted the striking

\* Corresponding author. Tel.: +44 20 7679 1031.

E-mail address: [e.crema@ucl.ac.uk](mailto:e.crema@ucl.ac.uk) (E.R. Crema).

resemblance of this process with the genetic theory of neutral evolution (Crow, 1986; Crow and Kimura, 1970; Kimura and Crow, 1964; Ewens, 1972), which provides mathematical foundations for determining how the frequency of cultural variants should be distributed for a given population size and innovation rate.

Thus, if we could calculate some statistical measure of how different cultural traits in a given archaeological assemblage are distributed and at the same time estimate what we should expect as a result of an unbiased cultural transmission, we should be able to obtain a quantitative insight on the evolutionary process behind the observed data. Neiman (1995) laid the foundation of this approach by introducing to the archaeological audience a diversity measure known as  $\theta$ . Its theoretical estimate, under neutral evolution/unbiased transmission, is  $2N_e\mu$ , where  $N_e$  is the effective population size, here defined as the number of entities actively contributing to the reproduction of cultural variants, and  $\mu$  is the rate of cultural innovation (see Neiman, 1995 for a detailed discussion on how this equation was obtained). Empirical estimates of  $\theta$  can be calculated in two ways. The first approach is to identify the value of  $\theta$  that satisfies the following equation (Ewens, 1972):

$$k = \sum_{i=0}^{s-1} \frac{\theta}{\theta + i} \quad (1)$$

where  $k$  is the observed number of cultural variants among a sample of size  $s$ . Since equation [1] does not have an analytical solution, one can obtain a maximum likelihood estimate which is referred to as  $t_E$ . The second empirical estimate of  $\theta$  is known as  $t_F$ , which is obtained by calculating the reciprocal of the sum of the squared proportions of each variant minus 1. If the empirical estimates  $t_E$  and  $t_F$  are equivalent to the theoretical expectation  $2N_e\mu$ , the observed frequency of cultural variants is indistinguishable from patterns expected by an unbiased cultural transmission, whilst dissimilarities would suggest an alternative process behind the observed pattern. The comparison between  $t_E$  (or  $t_F$ ) and  $2N_e\mu$  does not, however, provide a significance level. Two statistical tests can overcome this problem by using alternative summary statistics: the Ewens–Watterson test (Watterson, 1977, 1978), which compares theoretical and empirical estimates of the homogeneity statistic  $F$  (equivalent to the summed squares of the proportion of each variant); and Slatkin's exact test (Slatkin, 1994, 1996), which compares the observed frequency distribution of the cultural variants against all possible configurations expected by the Ewens sampling distribution for specific values of  $s$ , the sample size, and  $k$ , the number of variants (see Steele et al., 2010 for archaeological applications of both tests).

A number of archaeological studies have applied these techniques to investigate the observed variation in the frequency of cultural variants. Shennan and Wilkinson (2001) analysed the pottery assemblage of Linearbandkeramik (LBK) early Neolithic settlements in western Germany and compared  $t_F$  against six possible values of  $\theta$  (based on different combinations of two estimates of  $\mu$  and three estimates of  $N_e$ ). The result showed a strong divergence in most cases, leading the authors to reject the null hypothesis. Given that the observed diversity measure  $t_F$  was consistently higher than  $\theta$ , Shennan and Wilkinson suggested that the pattern of cultural transmission behind the observed archaeological record was affected by an anti-conformist bias, whereby variants with smaller frequencies have a higher probability of being copied compared to what we might expect in an unbiased transmission process. Kohler et al., (2004) used the same technique and examined the ceramic materials at Pajarito Plateau in New Mexico. Comparisons of the empirical estimates and the theoretical expectations of  $\theta$  indicated the opposite pattern to the one observed by Shennan and Wilkinson, leading the authors to postulate the presence of a conformist

transmission bias, whereby variants with higher frequencies have a higher probability of being copied compared to what we might expect with an unbiased cultural transmission.

These examples show how the biological theory of neutral evolution can be successfully applied to cultural domains, but at the same time highlight several issues. First, different forms of sampling bias affecting archaeological assemblages might lead to erroneous estimates in the relationship between observed and expected values of  $\theta$ . While the problem of sample size is tackled by equation [1], the problem of time averaging has been virtually ignored by most empirical studies. Archaeological assemblages are artificially defined and temporally varying sample blocks extracted from what is in fact a continuum. Other things being equal, assemblages associated with time blocks of longer duration are likely to have higher values of  $k$ , in turn determining higher estimates of  $t_E$ . A recent simulation study by Premo (2014) has shown that analyses based on the evaluation of  $\theta$  are prone to a type I error (i.e. an incorrect rejection of the null hypothesis) when the sample data is the result of a cumulated archaeological assemblage (see also Madsen, 2012). Alternative methods for detecting unbiased cultural transmission, such as the power-law model proposed by Bentley and colleagues (Bentley and Shennan, 2003; Bentley et al., 2004)<sup>1</sup> appear to be more robust in this case, although the effect of sample size and the duration of the assemblage formation are still relevant (Premo, 2014).

Second, direct computation of  $\theta$  requires reliable estimates of  $N_e$  and  $\mu$ . This is generally obtained from an inferential exercise based on external proxies (e.g. number of novel traits appearing in a new phase, number of residential features, etc.), but often requires the adoption of several alternative estimates to better incorporate competing scenarios and interpretations. This will ultimately lead to the generation of multiple variants of the same null hypotheses. For instance, Shennan and Wilkinson (2001) estimated two possible mutation rates and three possible effective population sizes, leading to a total of six version of the same null model. Given the wide range of possible values that can be generated by the same process of neutral evolution, under different effective population size and mutation rate, this approach might potentially lead to a type I error. Furthermore, conflicting results will require further investigation in support of specific null hypotheses.

Third, Steele et al. (2010) have shown how the failed rejection of the null hypothesis does not imply that the generative process was in fact the result of unbiased transmission (type II error). Their analyses of Hittite ceramic bowl types by means of Slatkin's exact test and the Ewens–Watterson homozygosity test failed to reject the null hypothesis. Nonetheless, further examination of the characteristics of the bowls showed how some features exhibited a significant positive correlation with the abundance ranking, which should not be expected under neutrality. Both the possible existence of multiple null models and the potential convergence in pattern of different generative process can be generalised to the well-known problem of equifinality (Von Bertalanffy, 1968:132, Premo, 2010 for a more recent discussion): a variety of different evolutionary mechanisms can potentially produce similar distributions in the frequencies of cultural variants.

Fourth, the model of neutral evolution discussed above assumes equilibrium conditions in the system of interest. This is not necessarily the case, as the underlying generative process could potentially

<sup>1</sup> Bentley and colleagues (Bentley and Shennan, 2003; Bentley et al., 2004) identified a power-law distribution in the frequency of cultural variants in both empirical and simulated data. Their simulation data suggest that the parameter  $\alpha$ , representing the negative slope of the power-law fit on a log–log plot, can be approximately estimated as  $0.1042 \times \ln(\mu N) + 1.48$ .

change its properties (e.g. a change in the innovation rate, or a transition from an unbiased to a conformist biased transmission), and the time to the new equilibrium will depend on the initial condition of the system. A recent paper by [Kandler and Shennan \(2013\)](#) has tackled this problem by shifting the focus to the number of cultural variants  $K(t_1, t_2)$  surviving the transition from a given time-period  $t_1$  to the subsequent time period  $t_2$ , rather than the frequency of cultural variants at a given time period  $t$ . Their mathematical model makes it possible to define the probability distribution of  $K(t_1, t_2)$  taking into consideration initial conditions of the system (the frequency of cultural variants at  $t_1$ ), variation in population size and mutation rate, and the temporal duration of the evolutionary process. They applied their model to a case study based on LBK ceramic assemblages from southwest Germany, showing how the empirically observed number of surviving variants was much higher than the one expected from neutral evolution. They further explored this divergence by developing a frequency-biased model (which includes both conformist and anti-conformist biases), showing that a transmission process with an anti-conformist bias has a higher fit to the observed data compared to the neutral model.

In order to overcome some of the issues raised by these previous works, we propose a simulation-based approach using Approximate Bayesian Computation (ABC, [Beaumont et al., 2002](#), [Csilléry et al., 2010](#)). This method offers several advantages that enable a more flexible approach for the analysis of changes in the frequency of cultural variants.

First, direct estimates of the model parameters, a key element in the direct evaluation of  $\theta$  (c.f. [Shennan and Wilkinson, 2001](#); [Kohler et al., 2004](#)) and in the non-equilibrium model proposed by [Kandler and Shennan \(2013\)](#), are substituted by a Bayesian inferential framework. Thus, instead of defining single or multiple estimates of each parameter (e.g. the innovation rate  $\mu$ ), we define a probability distribution known as a prior. This will formally define the range of values that we assume for a given parameter as well as the degree of belief we associate with each distinct value. The combination of this prior knowledge with the probability of obtaining the observed pattern for different parameter combinations of the proposed model, allows us to generate a posterior estimate of each parameter. Again, these will be probability distributions rather than point estimates, an update of the prior which incorporates the information from the empirical data as well as the specifics and the uncertainties of a given model.

Second, we abandon the hypothesis-testing approach and adopt an inferential framework based on multi-model selection ([Burnham and Anderson, 2002](#)). We thus formally integrate the notion that “all models are wrong” and seek to define the “best” model among a set of alternative candidates. The hypothesised model that has the highest fit to the empirical data and relies on the smallest number of assumptions (based on the principle of parsimony) will be selected. In the case of equifinality, two or more candidate models will be indistinguishable from a model selection perspective.

Third, following [Kandler and Shennan \(2013\)](#), we emphasize the importance of the non-equilibrium conditions of the system and hence measure differences in the frequencies of variants between different assemblages rather than looking at summary statistics (such as diversity indices) describing the frequencies observed at individual phases. The choice of summary statistic is undoubtedly a critical element in the selection between candidate models, as their sensitivity to the variation in the underlying process will differ. It should be noted that the choice in the present study has been partly driven by the specific research question of our case study, and that the proposed method can be adapted to any alternative measure of culture change.

Lastly, we tackle the problem of sample size and time averaging by integrating these processes in the simulation. More generally,

the key advantage of ABC resides in the fact that models can be defined as computational algorithms, enabling a flexibility that is harder to achieve with a pure equation-based solution.

The paper is structured as follows. Section 2 introduces three evolutionary models of cultural change that can be applied in a broad set of contexts; Section 3 presents a generic ABC-based workflow that can be applied for studies of cultural change; Section 4 illustrates our case study; and Section 5 concludes the paper by summarising our main points and highlighting the limits and the potentials of the proposed method.

## 2. Evolutionary models of cultural change

For the purpose of our paper we chose to explore three simple models of cultural change. Although the simulation-based approach allows us to integrate additional assumptions to incorporate a wide array of biases in the transmission process (see [Boyd and Richerson, 1985](#); [Henrich and McElreath, 2003](#)), we purposely chose to ignore biases favouring specific cultural variants based on their intrinsic properties (i.e. content-bias transmission), or biases derived from external cues unrelated to the variant of interest (i.e. model-based biased transmission). This choice was dictated by a preference for the most parsimonious and general models of cultural evolution as well as the lack of data concerning potential content and model biases (but see [Mesoudi and O'Brien, 2008a, 2008b](#)). We thus chose to explore an unbiased neutral model of cultural transmission and two different forms of frequency bias: anti-conformism and conformism.

### 2.1. Unbiased transmission (UB)

Unbiased cultural transmission (or neutral evolution) assumes that the probability of copying any given cultural trait is proportional to the frequency of that trait. Consequently, changes in the frequencies of variant will result from random events in the copying process (i.e. drift) and the introduction of novel variants through innovation. More formally, the probability  $\pi_i$  of copying a trait  $i$  in a given unit of time is defined by the following equation:

$$\pi_i = z \frac{m_i(1 - \mu)}{N} \quad (2)$$

where  $z$  is the frequency of cultural transmission,  $m_i$  is the existing instances of trait  $i$  (e.g. the number of potential teachers possessing the trait  $i$ ),  $\mu$  is the probability of introducing a new cultural variant (e.g. a new arrowhead type), and  $N$  is the number of potential “teachers” (i.e. the agents in the agents-based simulation), each displaying a single cultural variant. Notice that we do not consider  $N$  as the effective population size as we no longer follow the standard assumptions of neutral evolution (i.e. the Wright–Fisher model of reproduction). In fact, when  $z$ , the frequency of social learning, is smaller than 1, cultural transmission does not occur all the time for all individuals, and as a consequence the number of social learners will become smaller than the number of teachers.<sup>2</sup> This parameter will also portray overlapping generations (as individuals will engage in social learning at different time-steps<sup>3</sup>), and more importantly will allow us to calibrate the correspondence

<sup>2</sup> If an individual has a trait  $j$  with frequency  $m_j$ , the probability of losing such trait in a given time-step is  $\mu + z(N - m_j)N^{-1} - \mu z(N - m_j)N^{-1}$ . This combines the probability of innovating and the probability of copying another variant.

<sup>3</sup> It is worth noting that the copying is still between different generations, and hence this model is different from the so-called Moran model of reproduction which includes within generation copying ([Moran, 1958](#); see [Aoki et al., 2011](#) for its application in cultural evolution).

between simulation and real world time (e.g. with  $z = 0.1$ , assuming that a simulation time-step corresponds to a single year, social learning will occur on average once a decade).

## 2.2. Frequency biases: conformism (CB) and anti-conformism (AB)

equation [2] assumes that the frequency of a trait is directly proportional to its probability of being adopted, with the term  $\mu$  having the same effect for all traits. As discussed in Section 1, one of the most common alternatives to the neutral model is a frequency bias where either: 1) less common traits have higher chance of being adopted (anti-conformism); 2) or common traits have an increased probability of adoption (conformist bias) compared to an unbiased model. We can portray both processes with the following equation<sup>4</sup>:

$$\pi_i = z \frac{(1 - \mu)(m_i N^{-1})^{1-b}}{\sum_{j=1}^k (m_j N^{-1})^{1-b}} \quad (3)$$

Where  $k$  is the number of variants, and  $b$  is a parameter that controls both the direction and the magnitude of the bias. When  $b > 0$ , rare traits will have higher probability of adoption compared to those predicted by equation [2] (anti-conformism), and conversely when  $b < 0$ , common traits will have a higher  $\pi$  than the one expected from an unbiased cultural transmission (conformism).<sup>5</sup> Furthermore, the greater is the value  $|b|$ , the stronger is the effect of the frequency bias in both directions. Notice also that when  $b = 0$ , the model is mathematically equivalent to equation [2], making the unbiased cultural transmission a special case of frequency bias. As for equation [2], the parameter  $z$  captures the frequency of cultural transmission, whilst  $\mu$  indicates the probability of innovation.

## 2.3. Cultural dissimilarity

The four parameters  $N$ ,  $\mu$ ,  $z$ , and  $b$  described in equations [2] and [3] can portray a wide range of possible evolutionary dynamics detectable by a variety of summary statistics. Here we chose to use a measure of dissimilarity,  $D$ , which quantifies difference in the frequency of cultural variants between any two blocks of time  $t_1$  and  $t_2$ . The rationale of this choice is dictated by the assumption that evolutionary processes are characterised by temporal autocorrelation, with  $D(t_1, t_2)$ , the dissimilarity between the frequency of cultural variants at  $t_1$  and  $t_2$ , positively correlated with  $\Delta(t_1, t_2)$ , the temporal distance between  $t_1$  and  $t_2$ . Different conditions of the same evolutionary process, however, are expected to generate different relationships between  $D(t_1, t_2)$  and  $\Delta(t_1, t_2)$ . For example, high mutation rates is likely to determine a higher turnover rate (see Bentley et al., 2007), leading to higher values of  $D(t_1, t_2)$  for the same  $\Delta(t_1, t_2)$ . Here, we use the Morisita–Horn index (Morisita, 1959; Horn, 1966), a measure of dissimilarity given by the following equation:

$$D_{MH}(t_1, t_2) = 1 - \frac{2 \sum_{i=1}^k m_i(t_1) \cdot m_i(t_2)}{\left( \frac{\sum_{i=1}^k m_i(t_1)^2}{M(t_1)} + \frac{\sum_{i=1}^k m_i(t_2)^2}{M(t_2)} \right) M(t_1) M(t_2)} \quad (4)$$

where  $m_i(t_1)$  and  $m_i(t_2)$  are the counts of the variant  $i$  at time  $t_1$  and  $t_2$  and  $M(t_1)$  and  $M(t_2)$  are the total sample size for each time-blocks. equation [4] essentially returns a measure of overlap, expressed by a numerical index bounded between 0 (identical assemblages) and 1 (assemblages with no variants in common). Morisita–Horn dissimilarity is also independent of sample size (see Wolda, 1981), and hence well suited for archaeological contexts.

Fig. 1 shows the average and the 95% percentiles of the Morisita–Horn dissimilarity  $D_{MH}$  for values of  $\Delta$  between 50 and 2000 time-steps, obtained from a total of 1000 simulation runs of the three models proposed above. The parameter space portrayed in Fig. 1 shows a positive correlation in most scenarios, with some exceptions observed where  $b = -0.01$ ,  $z \geq 0.5$ , and  $N = 1000$ , when high levels of conformism enforce the fixation of few variants with no turn-over. Fig. 1 also shows the complex interplay between the four parameters and how different combinations can easily generate similar patterns, highlighting potential equifinality issues.

## 3. The Approximate Bayesian Computation (ABC) framework

We used an ABC framework to 1) determine probabilistic estimates of each of the three models described above and 2) select the “best” candidate model on the basis of its fit and complexity. Detailed discussion of the statistical principles and the broad inferential framework of ABC in biology can be found elsewhere (Csilléry et al., 2010 for a recent review; see also Kandler and Laland, 2013 for its theoretical application on cultural evolution studies and Bramanti et al., 2009 for an analysis of aDNA data to address the probability of genetic continuity between Mesolithic foragers and early Neolithic farmers in Europe); here we highlight its core algorithm and how this can be applied for our purpose.

### 3.1. Parameter estimation

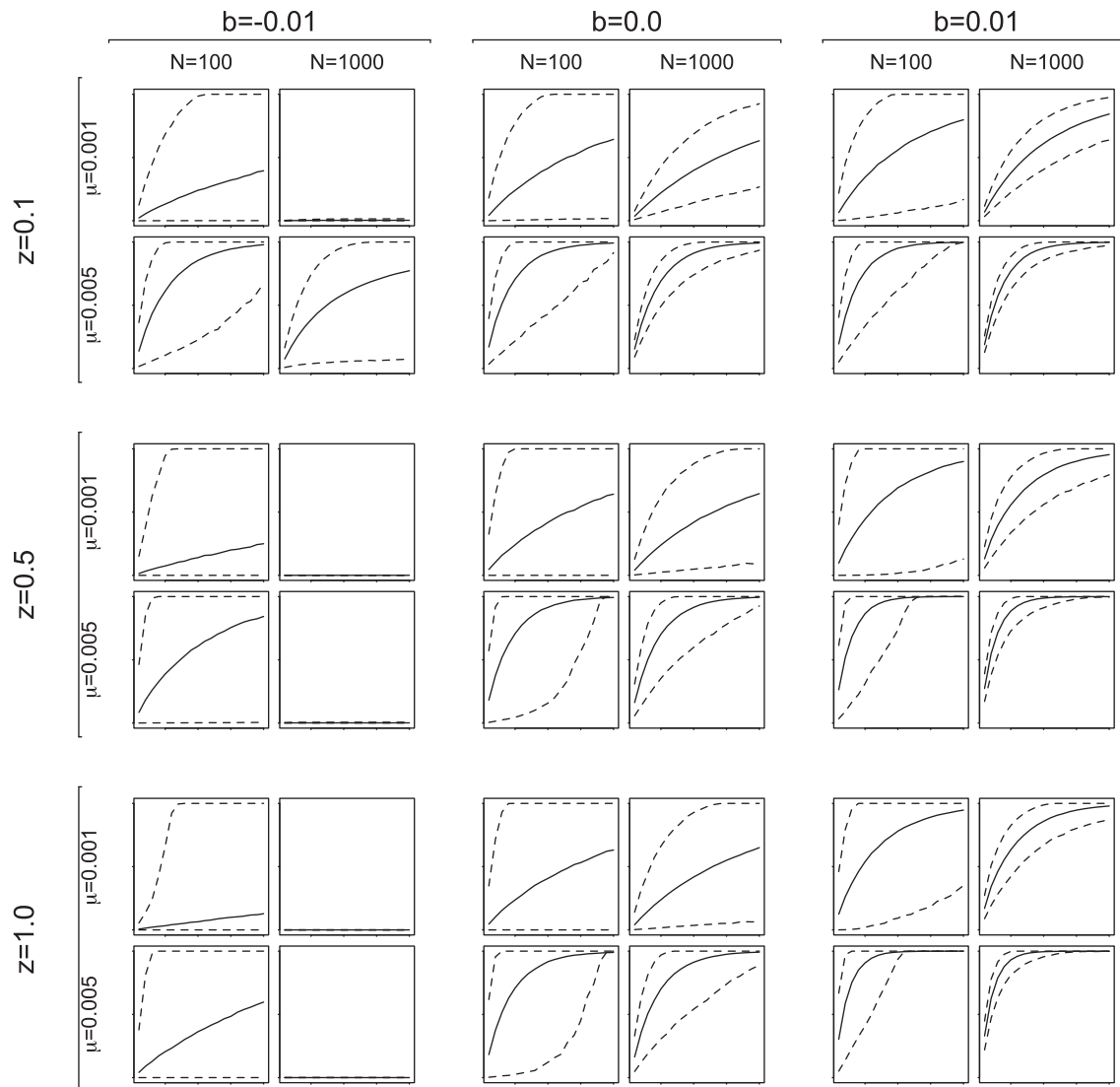
The parameter estimation process is rooted in the rejection algorithm originally developed in population genetics and can be summarised as follows:

- (1) Given observed archaeological data composed of  $T$  archaeological phases, each composed of a vector of occurrences of cultural variants, compute the dissimilarity index between each of the  $(T^2 - T)/2$  possible pairs. The resulting vector is the observed summary statistic  $S_{observed}$ .
- (2) For each hypothesised model define a prior probability distribution of its parameters (e.g. UB:  $N \sim$  Uniform probability distribution between 50 and 1000;  $\mu \sim$  Uniform probability distribution between  $5 \times 10^{-4}$  and  $10^{-2}$ ).
- (3) Sample  $n$  values from each parameter's prior distribution.
- (4) Execute the simulation and retrieve a vector of simulated summary statistics  $S_1, S_2, \dots, S_n$ .
- (5) Compute, for each of these, the Euclidean distances  $\delta_1, \delta_2, \dots, \delta_n$  to  $S_{observed}$ . This will give a measure of goodness-of-fit for each simulation run.
- (6) Select the proportion  $\tau$  (i.e. the “tolerance level”) of the simulations with the smallest  $\delta$  (i.e. the highest fit to the observed data).
- (7) Extrapolate the parameter values used to obtain the subset defined in (6).

<sup>4</sup> Models of conformist/anticonformists bias have been proposed by a number of scholars following the pioneer work by Boyd and Richerson (1985). Mesoudi and Lycett (2009) portray the bias as the selection of the single most (or least) frequent trait in the population, with a parameter representing the frequency by which this process occurs. Kandler and Shennan (2013) define the bias as a continuous function that is not limited to the selection of a single most common or rare variant. We followed the second approach as it encapsulates potential errors in detecting common/rare variants, but we also developed a simpler equation better suited for the application of ABC.

<sup>5</sup> The bias parameter is theoretically capable of portraying a high degree of variation in frequency bias, including extreme forms of anti-conformism where the correlation between  $\pi$  and  $m$  becomes negative (when  $b > 1$ ), hence rare traits becoming more likely to be adopted than frequent ones.





**Fig. 1.** Variation in the relationship between cultural dissimilarity and distance in time for different parameter values of  $N$  (population),  $\mu$  (innovation rate),  $z$  (probability individual engages in transmission process), and  $b$  (bias parameter – positive values = anti-conformism, negative values = conformism). Average values (solid lines) and 95% confidence (dashed line) obtained from 1000 simulations with 10,000 time-steps each.

The values obtained in the last step of this workflow provide a probabilistic estimate of the model parameters and can be effectively regarded as a posterior distribution that combines our prior belief (defined in point 2) with the knowledge obtained from the data (points 5–6).

### 3.2. Model comparison

The ABC framework offers a number of alternative algorithms for guiding the selection of the best candidate model. Here we present the most straightforward solution (but see [Francois and Laval, 2011](#) for more sophisticated techniques based on Deviance Information Criterion), although recent studies suggest caution in their interpretation ([Robert et al., 2011](#)).

The basic algorithm is an extension of the rejection method described above. In our case, the simulation outputs of each of the three competing models are simultaneously ranked on the basis of their Euclidean distance to the observed summary statistic. We then select a subset  $B$  of the best simulations, defined as the proportion  $\tau$  of the simulation with the closest distance to the

empirical summary statistic (i.e. the smallest  $\delta$ ). The proportion of each candidate model within the subset  $B$  can provide a crude estimate of which model has the best fit. These estimates can then be converted into Bayes factors ([Jeffreys, 1961](#)), a Bayesian version of the likelihood ratio test that can provide numerical support for the strength of evidence of one model over another.

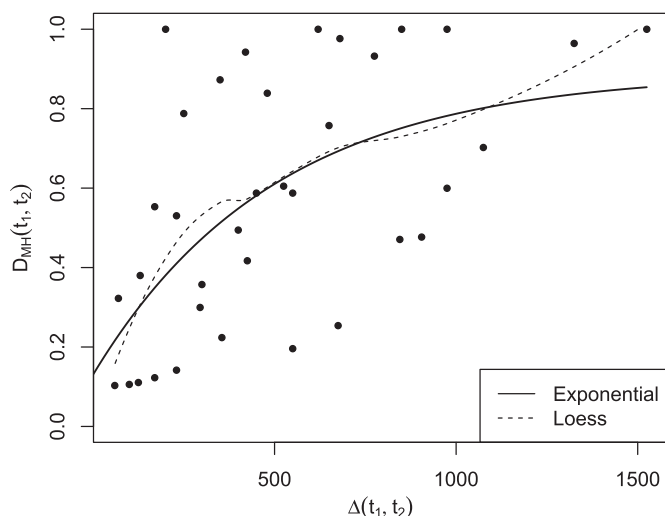
## 4. Case study

### 4.1. Materials and contexts

Our case study examines changes in the frequency of armature types from the Clairvaux and Chalais sites ([Pétrequin, 1986, 1989, 1997](#); [Pétrequin and Bailly, 2004](#)) in the Jura region of southeast France. The dataset (see [Electronic Supplement Table 1](#)) comprises a total of 280 arrowheads subdivided into nine chronologically distinct phases (I: 3700–3600 BC; II: 3450 BC; III: 3200 BC; IV: 3100 BC; V: 3050–3010 BC; VI: 3010–2930 BC; VII: 2850–2750 BC; VIII: 2750–2600 BC; IX: 2600–1650 BC) defined by [Saintot \(1998\)](#). For each phase, we counted the number of

arrowheads assigned to twenty distinct types, each defined as a unique combination of a binary presence-absence of morphological traits based on a paradigmatic classification (see [Electronic Supplement Table 2](#)) that excludes size and material. In this study we chose to analyse the arrowheads at the aggregated type scale of analysis instead of individual trait-level of analysis, for a closer comparison with Saintot's original analysis. Future studies will compare both analytical scales following [Edinborough 2005, 2008](#). [Fig. 2](#) shows the empirically observed relationship between  $D_{MH}(t_1, t_2)$  and  $\Delta(t_1, t_2)$ , the latter obtained as a distance between the median dates of each pair of phases. The scatter plot shows a clear positive correlation, with an exponential curve ( $R^2 = 0.35$ ) showing a slightly better fit compared to a linear relationship ( $R^2 = 0.32$ ). The overall pattern conforms to a general expectation of temporal autocorrelation (cf. [Fig. 1](#)), albeit with high levels of variations in  $D_{MH}$  for similar values of  $\Delta$ .

The choice of the case study was dictated by two sets of reasons. First, the comparatively small sample size is balanced by a robust chronological framework collated from dendrochronologically- and radiocarbon-dated sequences of superpositioned archaeological units, extracted from the remarkably well-preserved Neolithic pile-dwellings. This offers the rare opportunity to examine long-term cultural changes in arrowhead technology in a chronologically defined context. Second, in the original analysis conducted by [Saintot \(1998\)](#), patterns of morphological variation were ascribed to two episodes of broader regional cultural change, the first c.3200 BCE, marked by evidence of contact to the East and the second with the appearance of Bell Beaker material at c.2500 BCE ([Pétrequin, 1998; Saintot, 1998, 207](#)). These episodes correspond to the transitions between phases II and IV, and between phases VII–IX respectively. It follows that if Saintot's hypothesis is supported by the empirical data, we should expect these transitions to show more marked change in the frequencies of different cultural variants compared to other periods. Thus measures of dissimilarity such as the one described in equation [4] should exhibit higher values during these transitions, but we need to take into account differences in the duration of each phase and their temporal distances as well as the random nature of the copying mechanism, which can generate an array of different values under the same conditions (cf. [Fig. 1](#)). To integrate these elements in our re-evaluation of Saintot's hypothesis, we propose a three-step workflow, which we apply to examine the second period of cultural



**Fig. 2.** Observed relationship between distance in time ( $\Delta$ ) and Morisita–Horn dissimilarity ( $D_{MH}$ ). The solid line shows the fit for an exponential model ( $R^2 = 0.35$ ), while the dashed line is a loess fit with the smoothing parameter  $\alpha$  set to 0.75.

**Table 1**  
Model parameters, symbols, and priors.

Symbol	Description	Prior	Relevant models
$N$	Number of agents/effective population size	Uniform $\sim(50-1000)$	All
$\mu$	Rate of innovation	Uniform $\sim(10^{-4}-10^{-2})$	All
$z$	Rate of cultural transmission	Uniform $\sim(0.03-0.1)$	All
$b$	Frequency bias ( $b > 0$ : anti-conformist; $b < 0$ : conformist)	Uniform $\sim(-0.5-0; CB)$ ; Uniform $\sim(0-0.5; AB)$	AB & CB

change, corresponding to the transitions VII–VIII and VIII–IX.<sup>6</sup> We first estimate parameter values for the three different models of cultural transmission introduced in Section 2, we then identify the best model(s) given our data-set, and finally compare the empirically observed dissimilarities for these transitions against the ones predicted by the best model (s).

#### 4.2. Model implementation

We implemented our models by using an agent-based simulation written in R statistical language ([R Core Team, 2013](#)).<sup>7</sup> We explored three models: unbiased (UB), anti-conformist (AB), and conformist (CB) transmission. Each variant of the simulation has been conducted for 2050 time-steps, each corresponding to 1 year. We assumed that social learning occurs with a fairly slow rate, once or twice a generation (i.e.  $z$  between 1/30 and 1/10). This is broadly in line with ethnographic evidence on the amount of learning time required for specific techniques (cf. [Roux et al., 1995](#) on bead making) as well as on the skill development of bow-arrow skills ([Hill and Hurtado, 1996](#)). We also assumed that successful instances of transmission consist of a complete adoption of the all the morphological traits of a given arrowhead, essentially equating the unique combination of arrowhead components (the paradigmatic type) to the basic unit of analysis. Partial variations in the components were thus regarded as a form of innovation.

Prior estimates of the other three parameters were not dictated by any particular assumption, and hence their probability distributions can be regarded as uninformative priors (i.e. not subjectively elicited), although boundary values were partly informed by exploratory runs of the simulation. [Table 1](#) shows the priors of the parameters used in the hypothesised evolutionary models proposed here.

At time-step 1, we generated  $N$  agents, all possessing different integer values representing a cultural variant, in this case an arrowhead type. During the subsequent time-steps, each agent: 1) randomly adopted an existing trait in the population following one of the equations described in Section 2; and 2) generated, with probability  $\mu$ , a new cultural variant (i.e. a new arrowhead type), by randomly drawing a new integer value. [Table 1](#) summarises the relevant parameters for each of the three variants explored here. To avoid the idiosyncrasies derived from the initial conditions of the model, we started each simulation run with 30,000 “warm-up” iterations. Multiple experimental runs, using the combination of parameter runs with the slowest rate of cultural evolution (i.e. smallest mutation rate, highest population, and lowest transmission rate) showed that equilibrium is reached before this number (see [Electronic Supplement Fig. 1](#))

<sup>6</sup> We excluded the first period of major cultural change from our analysis given the extremely small sample size during phases I ( $n = 2$ ) and II ( $n = 3$ ).

<sup>7</sup> Source code and sample scripts are available in the online electronic supplement.

At the end of 30,000 + 2050 time steps we:

- 1) Discarded the output of the first 30,000 time steps;
- 2) Generated nine sets aggregating the distribution of cultural variants of the remaining 2050 time-steps following the observed length of the archaeological phases at Clairvaux and Chalais, discarding intervals where no archaeological phases were present.
- 3) Randomly sampled  $n$  cultural variants from each set, with  $n$  corresponding to the number of arrowheads recovered at each phase.
- 4) Calculated the frequency of each cultural variant for each of the nine sets/phases.
- 5) Computed the Morisita–Horn dissimilarity between all possible pairs of sets/phases

Each simulation thus portrayed specific cultural transmission processes as well as the archaeological uncertainties derived from sampling and time averaging, allowing a more exhaustive generative model of the archaeological assemblage at the Clairvaux and Chalais lake sites.

#### 4.3. Tackling sample size issues

Although the ABC framework is capable of integrating different sources of uncertainty in the posterior estimates of the parameters, the small sample size of our dataset might lead to some problems in the evaluation of the target summary statistics. Here we approach the problem by: 1) excluding the first two phases ( $n_I = 2$ ;  $n_{II} = 3$ ) from our analysis, and thus reducing the vector of target summary statistics from 36 to the 21 Morisita–Horn dissimilarities observed between the remaining seven assemblages (phases); and 2) using bootstrap estimates, rather than single point values of the observed summary statistics, for computing the distance between the simulated and empirical observed dissimilarity measures.

We achieved the second point by conducting 1000 bootstrap iterations for each phase by randomly sampling with replacement count values of each armature type. As a result we obtained a total of 21,000 Morisita–Horn dissimilarities. We then slightly modified the basic ABC workflow described above, by comparing each simulated vector of Morisita–Horn dissimilarities against different empirical Morisita–Horn dissimilarities that we randomly extracted from the bootstrap distribution. This ensured that both the parameter posterior and the model selection estimates incorporate the uncertainty of the target summary statistic.<sup>8</sup>

#### 4.4. Results

##### 4.4.1. Parameter inference

We conducted 100,000 simulations for each of the three hypothesised evolutionary models. Posterior distributions of each model have been extrapolated setting  $\tau$  to 0.01, equivalent to the 1000 simulations (out of 100,000) with the closest Euclidean distance to the observed summary statistic.

Fig. 3 shows the marginal posterior distribution of the four parameters for each of the three hypothesised models, and Table 2 shows the 95% credible region of each. Innovation rates show two distinct patterns, a strong peak at lower values (ca. 0.002–0.003) for the unbiased and anti-conformist models, and a wider and slightly bimodal posterior distribution for the conformist model. This bimodal shape is a good example of the complex interplay

between different parameters, with multiple combinations generating similar fit to the observed data (see online [Electronic Supplement Fig. 2–4](#)). Posterior estimates of the population size ( $N$ ), show a convergence of all models towards low values (50–100), albeit unbiased transmission exhibits a wider credible region compared to the frequency biased models. The frequency of transmission process ( $z$ ) shows instead high levels of uncertainty, although anti-conformist and unbiased models show some similarity to each other with slight peaks towards higher values, whilst conformist bias exhibits the opposite pattern. Both anti-conformist and conformist models show a peak towards 0 in the frequency bias parameter  $b$ , indicating that weak levels of frequency bias exhibit the highest fit to the observed data. However, it is worth noting that anti-conformist bias shows a wider credible region, (up to 0.4 for the 50% region) indicating that some levels of anti-conformist bias can also produce dissimilarities values similar to the observed ones.

##### 4.4.2. Model selection

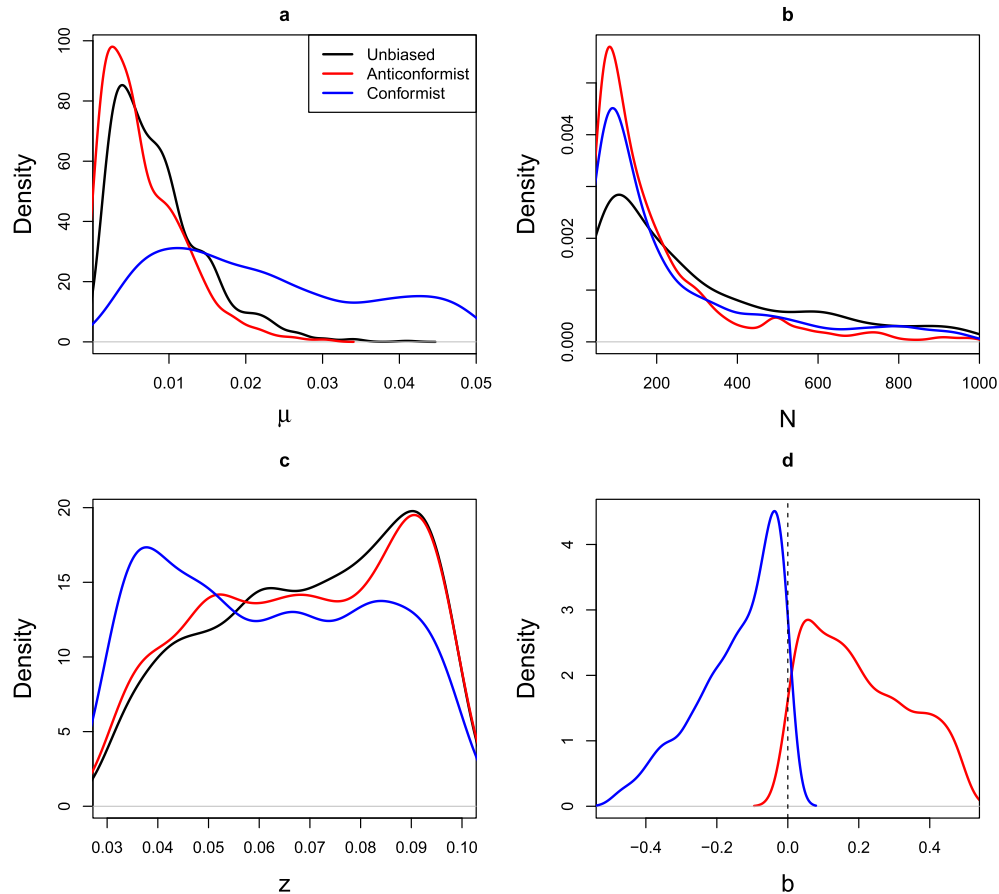
We compared our three models by computing an additional 100,000 runs for each using this time a parameter range derived from the posteriors obtained from the analysis discussed above. More specifically, we used uniform priors with the ranges defined by the 50% credible regions shown in Table 2.

Table 3 shows the result of the model comparison based on the standard rejection algorithm (with  $\tau = 0.01$ ) and the Bayes factor between all possible pairs. The results indicate that UB is the best model with a share of 47.5% of the 3000 simulations with the highest fit to the data. AB shows an equally high share of 43.9%, whilst only 8.6% of the best-fit simulations were from the CB model. One-to-one comparison based on Bayesian factors further confirm these values, with high support (i.e. Bayes Factors >5, see [Kass and Raftery, 1995](#) for interpretations of Bayes Factors) of UB and AB over CB, but no basis for choosing UB over AB (Bayes Factor = 1.08). In other words we can safely state that a transmission process based on conformism was highly unlikely, but the strong degree of equifinality between unbiased and anti-conformist transmission does not allow us to provide a single answer. However, it is worth highlighting again that unbiased transmission can be regarded as a special case of a frequency biased transmission with  $b = 0$ . In other words, UB is a point hypothesis, and its range of expected summary statistics, determined by the random nature of the copying process, substantially overlaps with the expected range of values expected from low frequency bias models.

Fig. 4 can illustrate this concept visually with a simulation. The solid line represents the probability density of the Morisita–Horn dissimilarity between two time-steps ( $t_1 = 30,000$ ,  $t_2 = 30,500$ ) obtained from 1000 simulation runs with  $N = 200$ ,  $\mu = 0.001$ ,  $b = 0$ ,  $z = 0.05$ . The dashed lines show instead the probability density of dissimilarities with three different values of  $b$ . The plot shows different degrees of overlap between the UB model and variants of the AB model as a function of  $b$ . When  $b = 0.01$ , the intrinsic variation in the simulation output (dictated by the random nature of the copying process) is larger than the difference between an unbiased and anti-conformist transmission process, making the two virtually almost indistinguishable. When  $b = 0.5$  the overlap is considerable smaller, but still present, suggesting in this case some degree of equifinality for dissimilarity values around 0.7.

The example illustrated in Fig. 4 prompts a more quantitative evaluation of the degree of overlap and equifinality between the proposed models. We achieved this by performing a leave-one-out cross validation making use of the existing simulations. This consists of iteratively selecting a random simulation output as a dummy target data, and executing the same model selection algorithm process adopted for the empirical data. The expectation is that the model that generated the dummy target has the highest

<sup>8</sup> Analyses with the standard workflow show qualitatively similar results, albeit with slightly narrower ranges in the posterior estimates.



**Fig. 3.** Marginal posterior distributions of the model parameters: a) mutation rate; b) population size; c) rate of cultural transmission; d) conformist/anti-conformist bias.

probability (retrieved from number of simulations) assigned among the target candidates. The result, known as a *confusion matrix* (Hastie et al., 2009, Csilléry et al., 2012), should have non-zero values only along the main diagonal in case of complete absence of misclassification. High values on non-diagonal elements will be the result of a large number of incorrect classifications, which in turn suggests a high probability that two or more competing models can generate similar pattern. In other words, the matrix provides a quantitative measure of the equifinality between competing models. Table 4 shows the confusion matrix of our three

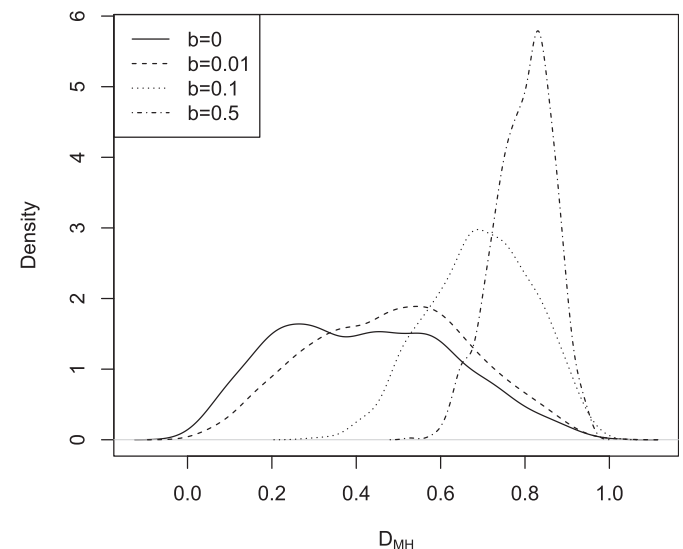
candidate models, showing how the 100 randomly sampled simulations of each model (rows) have been attributed (columns). The table confirms the presence of a strong degree of equifinality between UB and AB. This is however caused mostly by simulation runs sampled from the unbiased transmission being more

**Table 2**  
95% and 50% Bayesian credible region of the parameter posterior distribution.

	$\mu$	$N$	$z$	$b$
UB 95%	$11 \times 10^{-4} \sim 282 \times 10^{-4}$	52–970	0.031–0.099	NA
50%	$27 \times 10^{-4} \sim 163 \times 10^{-4}$	74–624	0.044–0.094	NA
AB 95%	$1.2 \times 10^{-4} \sim 255 \times 10^{-4}$	51–929	0.031–0.099	0.0025–0.4912
50%	$10 \times 10^{-4} \sim 148 \times 10^{-4}$	60–434	0.039–0.095	0.0278–0.4282
CB 95%	$19 \times 10^{-4} \sim 490 \times 10^{-4}$	50–924	0.030–0.099	–0.4676 – –0.0017
50%	$59 \times 10^{-4} \sim 429 \times 10^{-4}$	62–572	0.034–0.093	–0.3064 – –0.0127

**Table 3**  
ABC model selection: percentage of accepted simulations for the three candidate models and pairwise matrix of Bayes factors.

	% Accepted sim.	UB	AB	CB
UB	47.5%	1	1.08	5.49
AB	43.9%	0.92	1	5.08
CB	8.6%	0.18	0.19	1



**Fig. 4.** Probability density of Morisita–Horn dissimilarity between two time-steps ( $t_1 = 30,000$ ,  $t_2 = 30,500$ ) obtained from 1000 with  $N_e = 200$ ,  $\mu = 0.001$ , and different values of  $b$ .



**Table 4**

Confusion matrix with rows indicating the source of 100 randomly sampled simulations and columns their attribution based on ABC.

	UB	AB	CB
UB	37	57	6
AB	32	64	4
CB	10	18	72

frequently classified as anti-conformist bias (57 times out of 100) rather than UB itself (also 37 times out of 100). The other two models show slightly more distinguishable patterns: AB and CB were correctly classified 64% and 72% of the times respectively.

#### 4.4.3. Hypothesis testing and outliers

We can now determine whether we could identify significant outliers in the key transition phases at Chalain and Clairvaux, using as our null the three best models identified by ABC. The rationale of the approach proposed here is similar to the residual analysis of a linear regression. Inference on the model parameters has been based on the 21 empirically observed Morisita–Horn dissimilarities between all possible phases between III and IX. This however does not imply that the goodness of fit would be equal for each observed inter-phase summary statistic, as some transitions might have been characterised by a process distinct from the majority of the data.

Here we compare the observed Morisita–Horn dissimilarities between phases VII to VIII, phases VIII to IX, and VII to IX that have been interpreted as episodes of major cultural change linked with the appearance of the Bell Beaker culture at Chalain/Clairvaux. If this hypothesis is correct, we should expect these transitions to be outliers, with dissimilarities higher than those expected from the null model.

Fig. 5 shows the probability distribution of the Morisita–Horn dissimilarities expected from our two best models (obtained from the 1% best simulations retrieved by the ABC workflow), and the empirically observed values (shown as vertical dashed line with the bootstrap distribution as a histogram). In all cases the empirical estimates tend to be significantly smaller (one-sided  $p$ -values < 0.05), rather being larger, than the model expectations. This allows us to conclude that, given our two null models, there is no sufficient archaeological evidence supporting Saintot's hypothesis of a more marked change in the distribution of armature attributes with the appearance of the Bell Beaker culture. The results suggest instead a significant slowdown in the rate of cultural change compared to other phases examined at Chalain/Clairvaux.

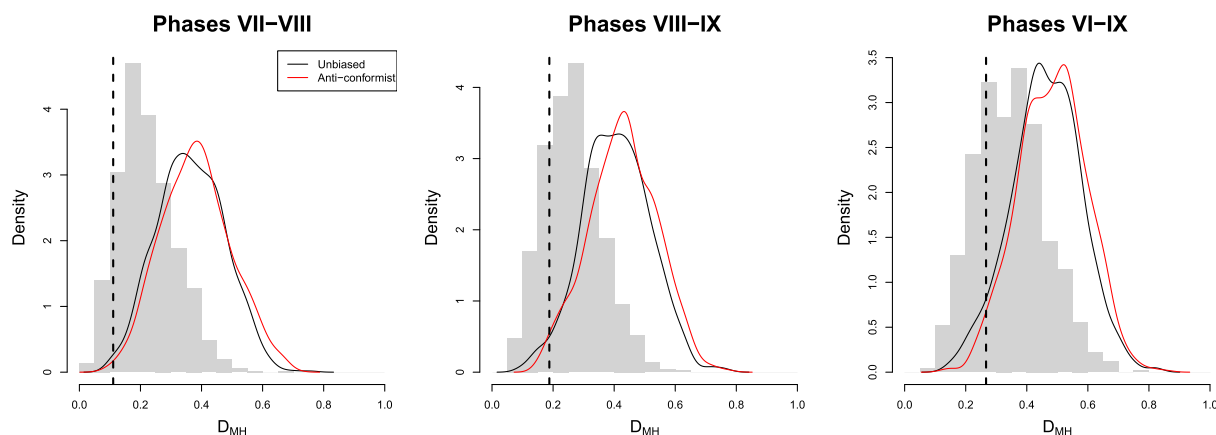
## 5. Discussion and conclusion

Inferring the evolutionary processes behind observed changes in the frequency of different cultural variants is difficult because of the comparatively poor quality of the archaeological data on the one hand and the large number of plausible alternative mechanisms of cultural transmission. The former limits the comparison between expectations derived from specific evolutionary models and the observed empirical data, while the latter introduces the problem of equifinality and multifinality (Premo, 2010).

We presented a solution based on the methodological framework of Approximate Bayesian Computation that can overcome some of these issues. Its simulation-based approach offers a flexible modelling environment where a variety of stochastic process can be integrated. Here we explored fairly simple evolutionary models, focussing part of our attention on issues pertaining to sample size and time averaging. However, more complex models can be explored within this framework, as long as the inferred generative process can be formally described with a simulation model. For example the assumption of the infinite allele model can be relaxed, introducing a parameter defining the probability of innovating a trait that has been already introduced in the past. This would better portrait cultural traits that are easily prone to be re-invented due to a limited design space.

One of the most appealing aspects of ABC is the possibility of shifting from a hypothesis-testing to a multi-model selection paradigm (Burnham and Anderson, 2002; see examples by Beheim and Bell, 2011; Towner et al., 2012; Eve and Crema, 2014). This enables the direct comparison of competing theories, offering the opportunity to identify which model provides the best fit to the data with the smallest number of assumptions. The multi-model paradigm offers also an objective account of how alternative models can actually generate similar patterns from different generative processes (cf. Table 3), providing a new way to tackle the problem of equifinality. We believe that these aspects make ABC an extremely useful tool for archaeological inquiry.

The application of ABC, however, has some limitations, partly intrinsic to its Bayesian foundations. Both the estimates of the posteriors and the model-selection process are confined by the boundaries defined by the user's assumptions (i.e. the priors or the set of candidate models), and hence results will vary accordingly. The two inferential exercises (parameter estimation and model selection) are also strongly related to each other. A large uninformative prior might provide a robust exploratory tool for identifying interesting regions within the parameter space of a single model,



**Fig. 5.** Expected Morisita–Horn dissimilarity for the UB (black solid line) and AB (red solid line) models, empirical observed values (dashed vertical line), and bootstrapped probability estimates (grey bars) for phases VII to VIII, VIII to IX, and VII to IX at Clairvaux and Chalain lake sites. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

but the dispersion of simulation runs will make it “weaker” in the comparative process. These fairly small drawbacks related to the implementation of ABC can be best approached by a balanced choice in the priors and candidate models informed by existing archaeological knowledge. Future archaeological applications of ABC might offer the necessary body of knowledge for facing these issues.

Our case study has offered some interesting points in this regard, and also in relation to previous works on neutral evolution discussed at the beginning of the paper. For example, we believe that summary statistics related to differences between assemblages are more useful than measures based on patterns within single assemblages. This enables us to assume a non-stationary model where a complex interplay of different generative processes occurs at different moments in time. We did not focus on identifying the best model for individual transitions (as in Kandler and Shennan, 2013), but instead concentrated on identifying an overall fit. The rationale in this case was to provide the highest number of summary statistics for the parameter inference to avoid equifinality issues (which we can expect to be high in the case of a small number of summary statistics and the choice of uninformative priors) and identify changes in the evolutionary process by looking at the model residuals (Section 4.4.3).

The high degree of equifinality identified by our study is partly due to the nested nature of the proposed models and partly to our conservative choice of using bootstrap estimates as target summary statistics. Although this analytical framework has undoubtedly increased the uncertainty in our output, we believe that integrating this information in a quantitative fashion is pivotal in archaeological research. This is not the first case of equifinality identified in evolutionary models of cultural transmission. Mesoudi and O'Brien (2008b) simulation-based study of Great Basin arrowheads indicated that the correlation between artefact attributes observed in their data could be explained by indirect bias as well as by neutral evolution and conformist bias. Perhaps future work exploring and possibly developing better statistics or a combination of targeted summary statistics might partly overcome this issue.

We did not explore transmission processes where the intrinsic properties of our cultural variants (e.g. performance of different armatures in hunting activities) can bias the social learning process, nor evolutionary dynamics related to individual learning. This choice was dictated by the lack of information concerning such properties in our dataset, and our, at least initial, preference for the most parsimonious and general models.

The results of the case study indicated that we do not have sufficient evidence for identifying a single best model describing the change in the frequency of armature types. Both unbiased and anti-conformist transmission models provide a good fit to the data, although the posterior estimates of the latter show a relatively low magnitude of this type of frequency bias. The results do not imply that alternative forms of social learning did not take place at Chalain and Clairvaux. Given the high degree of time-averaging, the fairly large temporal distances between some phases, and the relatively small sample size, detecting alternative modes of social learning requires a strong and stationary signal through-out the temporal scope of our analysis. Rapid changes between different forms of cultural transmission might also explain the observed pattern. This is also a cautionary tale for interpreting the rejection of the hypothesis proposed by Saintot. We did not find empirical evidence of higher dissimilarity during the appearance of the Bell Beaker culture around c.2500 BCE, and instead found significant support for a slowdown in the rate of cultural evolution, with the maintenance of similar frequencies in different armatures types. Our results are certainly valid only within the context of the best models amongst the candidates proposed here, and alternative

models might well show a different outcome. Nonetheless, we believe that our results are strong enough to suggest a reconsideration of previous hypotheses on the armature assemblages at Chalain and Clairvaux.

We believe our case study highlights both the limits and potential of this methodological framework. More generally we hope this work helps to foster the wider implementation and utility of more formal models in archaeology and their evaluation against the archaeological record.

## Acknowledgements

We thank the European Research Council for its Advanced Grant for project 249390, EUROEVOL, Cultural Evolution of Neolithic Europe, for enabling us to work on this paper. We are grateful to our three reviewers, Michael O'Brien, Luke Premo, and an anonymous reviewer who provided key insights, feedbacks, and critiques that helped improving the quality of the work. We are also grateful to Anne Kandler and Adrian Timpson for stimulating discussions on several key aspects of this work. All analyses have been conducted in R (R Core Team 2013), using the *vegan* (Oksanen et al., 2013) and *abc* (Csilléry et al., 2012) packages. We acknowledge the use of the UCL Legion High Performance Computing Facility (Legion@UCL), and associated support services, in the completion of this work.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jas.2014.07.014>.

## References

- Andersson, C., 2011. Paleolithic punctuations and equilibria: did retention rather than invention limit technological evolution? *PaleoAnthropology* 2011, 243–259.
- Aoki, K., Lehmann, L., Feldman, M.W., 2011. Rates of cultural change and patterns of cultural accumulation in stochastic models of social transmission. *Theor. Popul. Biol.* 79, 192–202.
- Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035.
- Beheim, B.A., Bell, A.V., 2011. Inheritance, ecology and the evolution of the canoes of east Oceania. *Proc. R. Soc. B* 278 (1721), 3089–3095.
- Bentley, R.A., Hahn, M.W., Shennan, S.J., 2004. Random drift and culture change. *Proc. R. Soc. B* 271 (1547), 1443–1450.
- Bentley, R.A., Lake, M.W., Shennan, S.J., 2005. Specialisation and wealth inequality in a model of a clustered economic network. *J. Archaeol. Sci.* 32, 1346–1356.
- Bentley, A., Shennan, S.J., 2003. Cultural transmission and stochastic network growth. *Am. Antiq.* 68, 459–485.
- Bentley, R.A., Lipo, C.P., Herzog, H.A., Hahn, M.W., 2007. Regular rates of popular culture change reflect random copying. *Evol. Hum. Behav.* 28, 151–158.
- Bettinger, R.L., Eerkens, J., 1999. Point typologies, cultural transmission, and the spread of bow-and-arrow technology in the prehistoric Great Basin. *Am. Antiq.* 64, 231–242.
- Boyd, R., Richerson, P.J., 1985. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.
- Bramanti, B., Thomas, M.G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M.N., Jankauskas, R., Kind, C.J., Lueth, F., Terberger, T., Hiller, J., Matsumura, S., Forster, P., Burger, J., 2009. Genetic discontinuity between local Hunter-Gatherers and Central Europe's first farmers. *Science* 326, 137–140.
- Burnham, K.P., Anderson, D.R., 2002. *Model Selection and Multi-model Inference: a Practical Information-theoretic Approach*, second ed. Springer, New York.
- Cavalli-Sforza, L.L., Feldman, M.W., 1981. *Cultural Transmission and Evolution: a Quantitative Approach*. Princeton University Press, Princeton.
- Cochrane, E., Lipo, C., 2010. Network and cladistic analyses of lapita decoration do not support branching evolution or regional population structure during colonization of western remote Oceania. *Philos. Trans. R. Soc. B* 365, 3889–3902.
- Collard, M., Shennan, S.J., Tehrani, J.J., 2006. Branching, blending, and the evolution of cultural similarities and differences among human populations. *Evol. Hum. Behav.* 27, 169–184.
- Conolly, J., College, S., Shennan, S., 2008. Founder effect, drift, and adaptive change in domestic crop use in early Neolithic Europe. *J. Archaeol. Sci.* 35, 2797–2804.

- Coward, F., Shennan, S., Colledge, S., Conolly, J., Collard, M., 2008. The spread of neolithic plant economies from the near east to Northwest Europe: a phylogenetic analysis. *J. Archaeol. Sci.* 35, 42–56.
- Crow, J.F., 1986. *Basic Concepts in Population, Quantitative, and Evolutionary Genetics*. W.H. Freeman, New York.
- Crow, J.F., Kimura, M., 1970. *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- Csilléry, K., Blum, M.G., Gaggiotti, O.E., Francois, O., 2010. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* 25, 410–418.
- Csilléry, K., Francois, O., Blum, M.G.B., 2012. abc: an R package for Approximate Bayesian Computation (ABC). *Methods Ecol. Evol.* 3, 475–479.
- Edinborough, K., 2005. Weapons of math's instruction: one thousand years of technological stasis in arrowheads of the Scandinavian Mesolithic. *Pap. Inst. Archaeol.* 16, 50–58.
- Edinborough, K., 2008. Population history, abrupt climate change and evolution of arrowhead technology in Mesolithic south Scandinavia. In: Shennan, S.J. (Ed.), *Pattern and Process in Cultural Evolution*. University of California Press, Berkeley CA, pp. 191–202.
- Eerkens, J.W., Lipo, C.P., 2007. Cultural transmission theory and the archaeological record: providing context to understanding variation and temporal changes in material culture. *J. Archaeol. Res.* 15, 239–274.
- Eve, S., Crema, E.R., 2014. A house with a View? Multi-model inference, visibility fields, and point process analysis of a Bronze Age settlement on Leskernick Hill (Cornwall, UK). *J. Archaeol. Sci.* 43, 267–277.
- Ewens, W., 1972. The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* 3, 87–112.
- Fitzhugh, B., 2001. Risk and invention in human technological evolution. *J. Anthropol. Archaeol.* 20, 125–167.
- Fort, J., 2012. Synthesis between demic and cultural diffusion in the Neolithic transition in Europe. *Proc. Natl. Acad. Sci. U. S. A.* 109.
- Francois, O., Laval, G., 2011. Deviance information criteria for model selection in approximate Bayesian computation. *Stat. Appl. Gen. Mol. Biol.* 10, 1–25.
- Hastie, T., Tibshirani, R., Friedman, J.H., 2009. *The Elements of Statistical Learning*, 2nd ed. Springer Verlag, New York.
- Henrich, J., McElreath, R., 2003. The evolution of cultural evolution. *Evol. Anthropol.* 12, 123–135.
- Henrich, J., 2004. Demography and cultural evolution: how adaptive cultural processes can produce Maladaptive Losses: the Tasmanian case. *Am. Antiquity* 69, 197–214.
- Hill, K., Hurtado, M., 1996. *Ache Life History. The Ecology and Demography of a Foraging People*. Aldine and de Gruyter, New York.
- Horn, H.S., 1966. Measurement of "Overlap" in comparative ecological studies. *Am. Nat.* 100, 419–424.
- Jeffreys, H., 1961. *Theory of Probability*, 3rd ed. Clarendon Press, Oxford.
- Kandler, A., Laland, K., 2013. Trade-offs between the strength of conformity and the number of conformists in variable environments. *J. Theor. Biol.* 332, 191–202.
- Kandler, A., Shennan, S., 2013. A non-equilibrium neutral model for analysing cultural change. *J. Theor. Biol.* 330, 18–25.
- Kass, R.E., Raftery, A.E., 1995. Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795.
- Kimura, M., Crow, J.F., 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49, 725–738.
- Kohler, T.A., VanBuskirk, S., Ruscavage-Barz, S., 2004. Vessels and villages: evidence for conformist transmission in early village aggregations on the Pajarito Plateau, New Mexico. *J. Anthropol. Archaeol.* 23, 100–118.
- Lake, M., Crema, E.R., 2012. The cultural evolution of adaptive-trait diversity when resources are uncertain and finite. *Adv. Complex Syst.* 19, 1150013.
- Lycett, S.J., von Cramon-Taubadel, N., 2014. Toward a "Quantitative genetic" approach to lithic variation. *J. Archaeol. Method Theory*. in press.
- MacDonald, D.H., 1998. Subsistence, sex, and cultural transmission in Folsom culture. *J. Anthropol. Archaeol.* 17, 217–239.
- Madsen, M.E., 2012. Neutral cultural transmission in time averaged archaeological assemblages. *arXiv* 1204.2043–2052 ([physics.soc-ph]).
- Mesoudi, A., Lycett, S.J., 2009. Random copying, frequency-dependent copying and culture change. *Evol. Hum. Behav.* 30, 41–48.
- Mesoudi, A., O'Brien, M.J., 2008a. The cultural transmission of great Basin projectile point technology I: an experimental simulation. *Am. Antiq.* 73, 3–28.
- Mesoudi, A., O'Brien, M.J., 2008b. The cultural transmission of great Basin projectile point technology II: an agent-based computer simulation. *Am. Antiq.* 73, 627–644.
- Moran, P.A.P., 1958. Random processes in genetics. *Math. Proc. Camb. Philos. Soc.* 54, 60–71.
- Morisita, M., 1959. Measuring of the dispersion and analysis of distribution patterns. *Memoires Fac. Sci. Kyushu Univ. Ser. Biol.* 2, 215–235.
- Neiman, F.D., 1995. Stylistic variation in evolutionary perspective: inferences from decorative diversity and interassemblage distance in Illinois Woodland ceramic assemblages. *Am. Antiq.* 60, 7–36.
- O'Brein, M.J., Lyman, R.L., 2003. *Cladistics and Archaeology*. University of Utah Press, Salt Lake City.
- O'Brien, M.J., Boulanger, M.T., Buchanan, B., Collard, M., Lyman, R.L., Darwent, J., 2014. Innovation and cultural transmission in the American Paleolithic: phylogenetic analysis of eastern Paleoindian projectile-point classes. *J. Anthropol. Archaeol.* 34, 110–119.
- Oksanen, J.F., Blanchet, G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.M., Wagner, H., 2013. *Vegan: Community Ecology Package*. R Package Version 2.0-10. <http://CRAN.R-project.org/package=vegan>.
- Peétrequin, P., 1986. Les sites littoraux néolithiques de Clairvaux-les-Lacs (Jura) –I- Problème matique géé;nérale. In: V exemple de la station III. de la Maison des Sciences de l'Homme, Paris.
- Peétrequin, P., 1989. Les Sites Littoraux Néolithiques de Clairvaux-les-Lacs (Jura) –II- Le Néolithique moyen. de la Maison des Sciences de l'Homme, Paris.
- Peétrequin, P., 1997. Les sites littoraux néolithiques de Clairvaux et de Chalain (Jura) –III- Chalain station 3, 3200–2900 av. J.-C. de la Maison des Sciences de l'Homme, Paris.
- Peétrequin, P., 1998. I. Les lacs de Chalain et de Clairvaux : dynamique éévolutive des styles céramiques et transferts de population. *Gall. Préhistoire* 40, 133–140.
- Peétrequin, P., Bailly, M., 2004. Lake-dwelling research in France. From climate to demography. In: Menotti, F. (Ed.), *Living on the Lake in Prehistoric Europe*. 150 Years of Lake Dwelling Research. Routledge, London, pp. 36–49.
- Powell, A., Shennan, S., Thomas, M.G., 2009. Late Pleistocene demography and the appearance of modern human behavior. *Science* 324, 1298–1301.
- Premo, L., 2010. Equifinality and explanation: the role of agent-based modeling in postpositivist archaeology. In: Costopoulos, A., Lake, M. (Eds.), *Simulating Change: Archaeology into the Twenty-first Century*. University of Utah Press, Salt Lake City, pp. 28–37.
- Premo, L.S., 2014. Cultural transmission and diversity in time-averaged assemblages. *Curr. Anthropol.* 55, 105–114.
- R Core Team, 2013. *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>.
- Richerson, P.J., Boyd, R., Bettinger, R.L., 2001. Was agriculture impossible during the Pleistocene but mandatory during the Holocene? a climate change hypothesis. *Am. Antiq.* 66, 387–411.
- Robert, C.P., Cornuet, J.-M., Marin, J.-M., Pillai, N.S., 2011. Lack of confidence in Approximate Bayesian Computation model choice. *Proc. Natl. Acad. Sci.* 108, 15112–15117.
- Roux, V., Bril, B., Dietrich, G., 1995. Skills and learning difficulties involved in stone knapping: the case of stone-bead knapping in Khambhat, India. *World Archaeol.* 27, 63–87.
- Saintot, S., 1998. Les Armatures de Flèches en Silex de Chalain et de Clairvaux In *Gallia préhistoire*, Tome 40, pp. 204–241.
- Shennan, S., 2001. Demography and cultural innovations: a model and its implications for the emergence of modern human culture. *Camb. Archaeol. J.* 11, 5–16.
- Shennan, S., Wilkinson, J.R., 2001. Ceramic style change and neutral evolution: a case study from Neolithic Europe. *Am. Antiq.* 66, 577–593.
- Shennan, S., 2011b. Property and wealth inequality as cultural niche construction. *Philos. Trans. R. Soc. B* 366, 918–926.
- Slatkin, M., 1994. An exact test for neutrality based on Ewens sampling distribution. *Genet. Res.* 64, 71–74.
- Slatkin, M., 1996. A correction to the exact test for neutrality based on Ewens sampling distribution. *Genet. Res.* 68, 259–260.
- Steele, J., 2009. Innovation diffusion and traveling waves. In: Shennan, S. (Ed.), *Pattern and Process in Cultural Evolution*. University of California Press, London, pp. 163–174.
- Steele, J., Glatz, C., Kandler, A., 2010. Ceramic diversity, random copying, and tests for selectivity in ceramic production. *J. Archaeol. Sci.* 37, 1348–1358.
- Towner, M.C., Grote, M.N., Venti, J., Borgerhoff Mulder, M., 2012. Cultural macro-evolution on neighbor Graphs: vertical and horizontal transmission among western North American societies. *Hum. Nat.* 23, 283–305.
- Von Bertalanffy, L., 1968. *General System Theory: Foundations, Development, Applications*. George Braziller, New York.
- Watterson, G.A., 1977. Heterosis or neutrality. *Genetics* 88, 789–814.
- Watterson, G.A., 1978. The homozygosity test of neutrality. *Genetics* 88, 405–417.
- Wolda, H., 1981. Similarity indices, sample size and diversity. *Oecologia* 50, 296–302.